

Spatial Analysis of Residential and Vacant Land using Mean Shift Clustering in Lhokseumawe City

Hafizh Al Kautsar Aidilof^{1*}, Munirul Ula², Rinaldi Mirsa³, Lidya Rosnita⁴, Marhaban Almaula⁵, Muji Budiman⁶

^{1,4,5,6}Informatics Engineering, Malikussaleh University, Lhokseumawe, Indonesia

Jl. Kampus Unimal Bukit Indah, Blang Pulo, Kec. Muara Satu, Kota Lhokseumawe,

²Master of Information Technology, Malikussaleh University, Lhokseumawe, Indonesia

³Architecture, Malikussaleh University, Lhokseumawe, Indonesia

¹hafizh@unimal.ac.id, ²munirul@unimal.ac.id, ³rinaldi@unimal.ac.id, ⁴lidya@unimal.ac.id, ⁵marhaban253@gmail.com



*Corresponding Author

Article History:

Submitted: 08-12-2025

Accepted: 17-12-2025

Published: 01-04-2026

Keywords:

clustering analysis;
Lhokseumawe; mean shift
algorithm; settlement patterns;
spatial planning; vacant land
utilization.

Brilliance: Research of

Artificial Intelligence is licensed
under a Creative Commons
Attribution-NonCommercial 4.0
International (CC BY-NC 4.0).

ABSTRACT

Uncontrolled urban growth in medium-sized Indonesian cities like Lhokseumawe has created spatial disparities through uneven settlement distribution and numerous underutilized vacant lands, demanding adaptive and data-driven planning approaches. This study aims to analyze spatial patterns of settlement and vacant land distribution in Lhokseumawe City by implementing the Mean Shift Clustering algorithm and to develop spatial planning recommendations supporting sustainable development. The research employs a quantitative approach using spatial data from OpenStreetMap (OSM), Sentinel-2 satellite imagery (for NDVI calculation), and Detailed Spatial Planning (RDTR) data. The non-parametric, density-based Mean Shift Clustering algorithm was applied to adaptively group areas without predetermined cluster numbers, analyzing both city-wide and Blang Mangat District levels. Results show Mean Shift successfully identified 62 spatial clusters across Lhokseumawe City and 28 clusters (OSM-based) plus 37 clusters (RDTR-based) in Blang Mangat District. Analysis revealed linear-transitional urban development patterns from the west (dense urban core) to northeast (transition areas and vacant land), identifying zones of dense settlements, transition zones potential for infill development, and vacant land suitable for green spaces or planned development. Findings also revealed discrepancies between factual spatial patterns from clustering and RDTR zoning plans. The study concludes that Mean Shift Clustering effectively reveals natural spatial structures of settlements and vacant land, providing credible data for spatial policy revision with main recommendations focusing on vertical development, transportation system development along main corridors, and conservation area designation in eastern and northern zones to support Lhokseumawe's sustainable urban planning.

INTRODUCTION

Uncontrolled urban growth is a common phenomenon in many medium-sized cities in Indonesia, including Lhokseumawe. This phenomenon is often characterized by uneven settlement distribution and a significant amount of underutilized vacant land. Such spatial imbalances have the potential to lead to land use inefficiencies, increased infrastructure costs, and vulnerability to environmental disasters and social problems. In the context of regional planning, evaluating development conformity with the Detailed Spatial Plan (RDTR) is crucial to ensuring sustainable urban development Asri et al., (2023). Furthermore, spatial patterns are often formed not only through physical adaptation but also as a manifestation of the community's socio-economic dynamics and cultural values (Haerani & Oktavia, 2023)

Previous research has demonstrated the effectiveness of spatial data mining approaches in analyzing regional patterns. Classical clustering methods, such as K-Means, have been widely used, for instance, in mapping landslide-prone zones Suwardi, (2020) and monitoring crime distribution Hamsinar et al., (2024). However, methods like K-Means have a primary limitation: they require the pre-determination of the number of clusters (k), which is often subjective and difficult to define for complex, irregular spatial data. Additionally, spatial analysis has been utilized for various health and social purposes, such as blood transfusion mapping Fitria et al., (2024) and dengue fever distribution analysis (Tahun et al., 2021) highlighting the importance of location-based data grouping accuracy.

To overcome the limitations of algorithms requiring cluster initialization, the Mean Shift Clustering algorithm serves as a robust alternative. This algorithm is non-parametric and density-based, allowing it to automatically identify clusters based on data density without the need to specify the number of clusters beforehand. The reliability of this method has been proven across various domains, ranging from water quality monitoring (Al et al., 2025) waste



management zoning Musaffak et al., (2025), electricity distribution clustering Utari et al., (2024) to non-cash food aid distribution Haerani & Oktavia, (2023) and fire department customer grouping (Caesar, 2023)

Despite the proven effectiveness of Mean Shift Clustering, its application for simultaneously mapping the spatial patterns of settlements and vacant land within the context of urban planning in Lhokseumawe remains limited. Based on these phenomena and research gaps, this study aims to analyze the spatial distribution of settlements and vacant land in Lhokseumawe City using the Mean Shift Clustering approach. This research is expected to generate accurate, data-driven information to support sustainable urban planning and serve as an evaluation material for the implementation of the existing RDTR.

LITERATURE REVIEW

The analysis of spatial patterns of settlements and vacant land is a growing field of research, driven by advances in geospatial technology. Settlement patterns are not formed randomly; they are influenced by various physical and non-physical factors. Previous studies, such as those on the settlement patterns of the Samudera Pasai Sultanate, indicate that spatial configuration is strongly influenced by historical and cultural values Mirsa et al., (2021). A similar finding was observed in Laweyan, where closed settlement patterns were formed to support the community's batik production activities Mirsa & Soetomo, (2016). In a modern context, these patterns must align with formal planning instruments like the RDTR to prevent deviations in land utilization (Asri et al., 2023)

In the technical realm, clustering algorithms are vital tools for identifying hidden structures within spatial data. The K-Means method is often a popular choice due to its simplicity, as applied in landslide zoning mapping Suwardi, (2020). and crime distribution analysis Hamsinar et al., (2024) However, the reliance on determining the initial k value often reduces its flexibility in handling dynamic, real-world data.

As an alternative, Mean Shift Clustering offers a more adaptive approach. This algorithm works by shifting data points toward the highest density (mode), allowing clusters to form naturally according to data density. The effectiveness of this method has been confirmed in several recent studies (Caesar, 2023) successfully implemented it to group fire retribution customers, while Aidilof et al., (2025) used it for water quality monitoring data clustering with precise results. The advantage of Mean Shift in handling data without assuming cluster shapes is also evident in research on non-cash food aid distribution Haerani & Oktavia, (2023) and electricity distribution data grouping Utari et al., (2024) This study will adopt the strengths of this algorithm to dissect the complexity of settlement and vacant land patterns in Lhokseumawe, providing a novel perspective compared to conventional methods, while building upon recent advances in spatial clustering for urban applications Gui et al., (2025), Comparative studies of spatial clustering algorithms have demonstrated that density-based methods like Mean Shift are particularly effective for urban land use analysis where natural cluster structures are unknown (Wihasti et al., 2025)

METHOD

This research was conducted through a systematic series of stages, beginning with a literature study and problem identification that comprehensively reviewed Mean Shift clustering, spatial analysis principles, and sustainable urban planning, alongside an analysis of city spatial planning regulations. The selection of Mean Shift Clustering was based on its proven effectiveness in handling spatial data with unknown cluster numbers and its successful applications in various urban studies (Radfar et al., 2025)

Subsequently, data collection and processing were undertaken by acquiring satellite imagery from Google Earth, Sentinel-2, and Landsat 8, which were supplemented by geospatial data from OpenStreetMap and the Detailed Spatial Plan (RDTR) obtained from the Public Works Department of Lhokseumawe City. The experimental design involved adjusting the spatial analysis scale and determining Mean Shift parameters, specifically optimizing bandwidth through cross-validation. Following this, the Mean Shift Clustering algorithm was implemented using Python with the scikit-learn library, where the results were spatially validated through reference maps and limited ground truth, and further evaluated using the silhouette score as a cluster validity index. The study then proceeded to spatial pattern analysis to identify and map clustering results through overlay analysis with the Regional Spatial Plan (RTRW) and road network maps. Finally, the findings were interpreted to formulate recommendations for sustainable urban planning, focusing on the analysis of settlement development potential and vacant land revitalization

The research utilized multiple data sources and parameters as shown in Table 1.

Table 1. Parameters used

No.	Parameter	Description
1	Geospatial Coordinates (X, Y)	Location points of each pixel in UTM or WGS84 coordinate systems
2	Building Density/Pixel	Calculated based on number of building pixels within specific radius
3	Land Cover Types	Classified from satellite imagery (settlement, vacant land, vegetation, water)
4	NDVI (Normalized Difference Vegetation Index)	Shows vegetation coverage - low values may indicate vacant land
5	Distance to Main Roads/Public Facilities	Used to assess development accessibility potential
6	Zoning in RTRW	Used as reference for clustering results validation in sustainable planning

Mean Shift Clustering is a non-parametric algorithm that determines clusters based on data density without requiring predetermined cluster numbers. The algorithm works through the following mathematical formulation.

Step 1 : Kernel Function and Bandwidth Determination

The Gaussian kernel function regulates the neighbor search range:

$$K(x) = \exp\left(-\frac{|x|^2}{2h^2}\right) \quad (1)$$

where h is the bandwidth that affects the distance or radius from the cluster center when calculating the average displacement. The Gaussian or Epanechnikov function (φ) allows assigning a weight to each data point according to its distance from the cluster center.

Step 2 : Initial Mass Centers Initialization

The initial mass centers that will become centroids are initialized as:

$$m^{(0)} = x_i \quad (2)$$

Step 3: New Mass Center Calculation

For each centroid, recalculate the new mass center point using the Mean Shift formula:

$$m^{(t+1)} = \frac{\sum_{i=1}^n K\left(\frac{m^{(t)} - x_i}{h}\right) x_i}{\sum_{i=1}^n K\left(\frac{m^{(t)} - x_i}{h}\right)} \quad (3)$$

The calculation of the new center of mass ($m^{(t+1)}$) is calculated based on the total of the kernel function ($K\left(\frac{m^{(t)} - x_i}{h}\right)$) divided by the total of the kernel function values ($K\left(\frac{m^{(t)} - x_i}{h}\right)$) within the given kernel radius, where $m^{(t)}$ is the current center of mass and x_i is the data point within the kernel radius.

Step 4 : Convergence Criteria

Repeat the new mass center calculation for each centroid until convergence:

$$\left\|m^{(t+1)} - m^{(t)}\right\| < \varepsilon \quad (4)$$

RESULT

The implementation of Mean Shift Clustering algorithm successfully identified spatial patterns of settlements and vacant land in Lhokseumawe City. Analysis results demonstrate that the algorithm formed 62 spatial clusters representing variations in settlement density and vacant land across the entire city area, as shown in Figure 1.

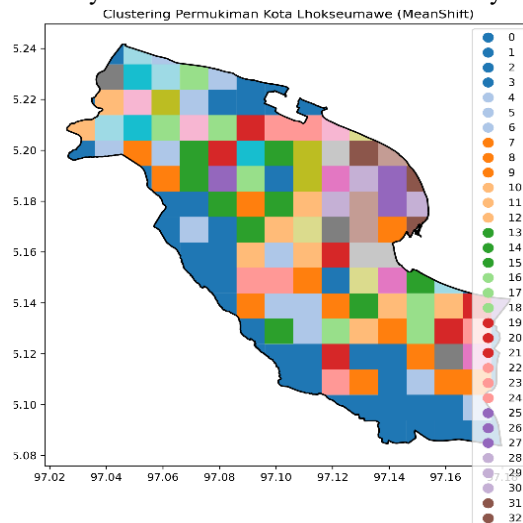


Fig 1. Results of Clustering of Lhokseumawe City Area using OSM

Detailed analysis in Blang Mangat District yielded 28 clusters when using OpenStreetMap data (Figure 2) and 3

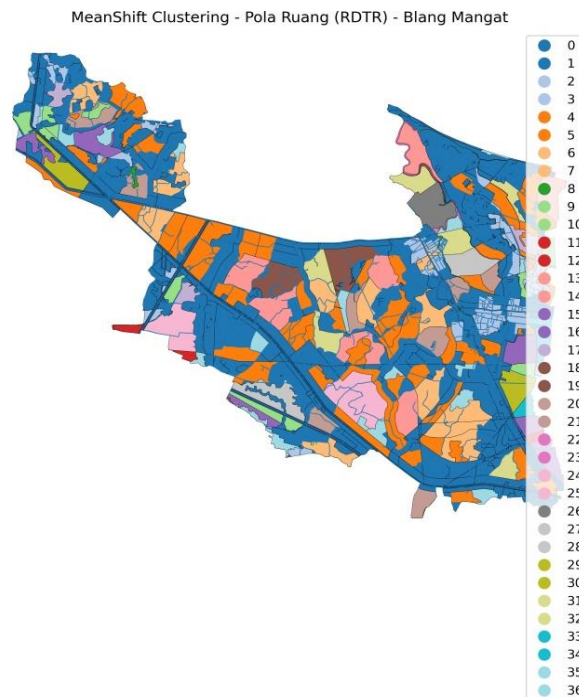


Fig 2. Results of Clustering of Blang Mangat District Area using OSM

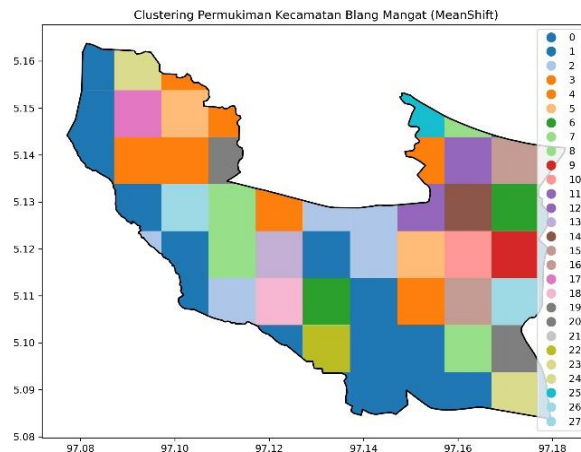


Fig 3. Results of Clustering of Blang Mangat District Area using RDTR

Identified spatial patterns reveal clear distribution distinctions between city core areas, transition zones, and suburban regions. The city center (Banda Sakti District) is dominated by clusters with high building density, while suburban areas show clusters dominated by vacant land and vegetation.

Table 2. RDTR Layers Used in Spatial Analysis

No	Layer	Description
1	_1173_5RD_AR_PR_RDTRLHOKSEUMAWE_2021	Spatial Pattern
2	_1173_5RD_LN_SR_TRANSPORTASI_RDTRLHOKSEUMAWE_2021	Road Network
3	_1173_5RD_PT_SR_PUSATPELAYANAN_LHOKSEUMAWE_2021	Service Center Point
4	Badan_Air	As an indicator
5	Pola_LP2B	Agricultural Land

The five RDTR layers listed in Table 2 provided the essential spatial context for interpreting the Mean Shift clustering results. These layers were not used as direct input features for the clustering algorithm (which relied on the parameters in Table 1), but were instead integrated through overlay analysis to add semantic meaning to the identified clusters.

The Spatial Pattern layer served as the primary reference to classify clusters into formal land-use categories (e.g., residential, mixed-use, conservation). Clusters overlapping with zones designated for settlements in the RDTR were flagged for conformity assessment.

The Road Network and Service Center Point layers were used to analyze cluster accessibility and functional centrality. This analysis explained why certain high-density clusters formed around major intersections or public facilities, revealing a discrepancy between the planned service hierarchy and the actual settlement patterns. The Water Body and Agricultural Land (LP2B) layers acted as constraining factors. Clusters identified in or adjacent to these zones highlighted areas where physical and regulatory limits naturally restrict urban expansion, pinpointing regions where the clustering results suggest a need for strict development control or green infrastructure planning.

In summary, the integration of these RDTR layers transformed the abstract numerical clusters into spatially intelligible planning units. This allowed for a direct comparison between the de facto patterns revealed by Mean Shift clustering and the de jure plans encoded in the RDTR, forming the basis for the policy discussions in the following section.

$$S = \frac{b-a}{\max(a,b)} \quad (5)$$

where S represents the silhouette score, a denotes the average intra-cluster distance, and b represents the average distance to the nearest cluster. The obtained silhouette score indicates good clustering quality, although variations were observed based on the bandwidth parameters used.

DISCUSSION

The difference in cluster numbers between data sources indicates the algorithm's sensitivity to input data characteristics. The success of Mean Shift Clustering in identifying clusters without predetermined numbers proves this algorithm's effectiveness for complex spatial data analysis. This finding aligns with research by Musaffak et al., (2025), who also successfully applied Mean Shift for area clustering based on waste management. The varying clustering results between OSM data and RDTR data confirm findings by Aidilof et al., (2025) that input data quality and characteristics significantly influence clustering outcomes. This difference reflects variations in accuracy and detail of information between crowdsourced data (OSM) and official government data (RDTR).



The identified spatial patterns provide an empirical basis for spatial planning, where high-density clusters in city centers indicate revitalization needs, while vacant land clusters in suburbs hold potential for planned area development. This approach is consistent with research emphasizing the importance of spatial data integration in RDTR evaluation (Asri et al., 2023)

The validation results are consistent with previous studies which also found high accuracy in Mean Shift application for spatial data clustering. The spatial patterns identified in this study provide practical implications for urban planning in medium-sized cities. High-density clusters in city centers suggest areas requiring infrastructure reinforcement and revitalization, while vacant land clusters in peripheral areas represent opportunities for sustainable urban expansion and green space development

This study has several limitations that should be acknowledged. The clustering results are highly dependent on data quality and resolution, particularly the variations between crowdsourced OSM data and official RDTR data. The Mean Shift algorithm's sensitivity to bandwidth parameter selection introduces subjectivity in the clustering process, despite optimization efforts through cross-validation. Additionally, the focus on physical spatial patterns without comprehensive integration of socio-economic factors may limit the holistic understanding of urban dynamics. Future studies should address these limitations by incorporating multi-source data validation and integrating socio-economic variables for more robust urban pattern analysis.

CONCLUSION

This research demonstrates that Mean Shift Clustering effectively identifies natural spatial patterns of settlements and vacant land in Lhokseumawe City without requiring predetermined cluster numbers. The algorithm successfully formed 62 spatial clusters across the city, with detailed analysis in Blang Mangat District revealing 28 clusters using OpenStreetMap data and 37 clusters using RDTR data. This variation highlights the algorithm's sensitivity to different data sources while maintaining robust pattern recognition capabilities. The research provides valuable insights for sustainable urban planning by clearly distinguishing between high-density settlement areas in city centers and vacant land clusters in suburban regions, offering empirical evidence to support targeted urban development strategies.

The research contributes significantly to urban planning practice through data-driven decision making, where clustering results provide quantitative basis for spatial planning policies and resource allocation. The identification of vacant land clusters enables strategic planning for green spaces and controlled urban expansion, supporting sustainable development goals. Furthermore, the approach successfully bridges data science, geospatial analysis, and urban planning disciplines, while collaboration with local government ensures direct applicability of research findings in practical settings.

Several limitations were identified during the research, including dependency on data quality and availability from various sources, sensitivity of clustering results to bandwidth parameter selection in Mean Shift algorithm, limited ground truth validation due to resource constraints, and focus on physical spatial patterns without comprehensive socio-economic factors. These limitations highlight areas for improvement in future research.

Based on the research findings and limitations, several recommendations are proposed for further study. Methodological enhancement through comparative studies using alternative clustering algorithms would help evaluate relative performance and accuracy in spatial pattern identification. Parameter optimization through automated bandwidth selection methods could improve consistency and reduce subjectivity in parameter tuning. Expanded data integration incorporating socio-economic variables, population density data, infrastructure availability, and environmental factors would create more comprehensive urban pattern analysis. Temporal analysis through longitudinal studies would track spatial pattern evolution over time and assess urban development dynamics. Technology development creating web-based geographic information systems with interactive visualization capabilities would enhance accessibility for stakeholders and support participatory planning processes. Finally, strengthened collaboration with government agencies would ensure research findings are effectively integrated into spatial planning policies and sustainable development frameworks, maximizing the practical impact of the research outcomes.

REFERENCES

- Al, H., Aidilof, K., Rosnita, L., & Ikhwan, M. (2025). *Clustering of Data Monitoring Water Quality Using Mean-Shift Clustering Method*. 6(1), 764–768. <https://doi.org/10.30596/jcositte.v6i1.22390>
- Asri, A., Mirsa, R., & Saputra, E. (2023). Evaluasi Pembangunan Kabupaten Pidie berdasarkan Rencana Detail Tata Ruang (RDTR)[Studi Kasus Kecamatan Pidie]. *Jurnal Rekayasa Teknik Dan Teknologi*, 7(2), 91–96.
- Caesar, R. A. (2023). *Implementasi Mean Shift Clustering dalam Mengelompokkan Pelanggan Retribusi Alat Pemadam Kebakaran pada Dinas Pemadam Kebakaran dan Penanggulangan Bencana Kota Palembang* [Politeknik Negeri Sriwijaya]. <https://eprints.polsri.ac.id/14845/>
- Fitria, N. A., Rangkuti, Y. M., Muslim, I., & Karo, K. (2024). Spasial Clustering Untuk Memetakan Transfusi Darah 25. *Jurnal JUPITER*, 16(1), 25–36.
- Gui, B., Sam, L., Bhardwaj, A., Gómez, D. S., Peñaloza, G., Buchroithner, M. F., Green, D. R., Gui, B., Sam, L.,

- Bhardwaj, A., Gómez, D. S., Peñaloza, G., Buchroithner, M. F., & Optimizing, D. R. G. (2025). Optimizing object-based agricultural and ecological land use classification via multi-band fusion and seasonal spectral differences. *International Journal of Remote Sensing*, 46(16), 6157–6188. <https://doi.org/10.1080/01431161.2025.2532837>
- Haerani, E., & Oktavia, L. (2023). Penerapan Algoritma Mean-Shift Pada Clustering Penerimaan Bantuan Pangan Non Tunai. 4(4), 1019–1027. <https://doi.org/10.47065/josyc.v4i4.3876>
- Hamsinar, H., Manarfā, W. O. R. A. U., & Salam, A. P. C. (2024). Metode K-Means Clustering Dan Analisis Spasial Untuk Monitoring Sebaran Kriminalitas. *Jurnal Informatika*, 13(2), 91–100. <https://doi.org/10.55340/jiu.v13i2.2151>
- Mirsa, R., Saputra, E., & Farhana, I. (2021). *Space Pattern of Samudera Pasai Sultanate*. 1(2), 94–103.
- Mirsa, R., & Soetomo, S. (2016). *Spatial Settlement Pattern in Laweyan as Supportive Batik Production Activities*. 48(1).
- Musaffak, A. L., Hindrayani, K. M., & Idhom, M. (2025). Penerapan Metode Mean Shift Clustering Untuk Mengelompokkan Wilayah Berdasarkan Pengelolaan Sampah. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4790–4797.
- Radfar, S., Koosha, H., Gholami, A., & Amindoust, A. (2025). *Unveiling Spatiotemporal Patterns in Public Transportation : A Smart Card Data Analysis*. <https://doi.org/10.22059/cej.2025.385897.2189>
- Suwardi, A. M. (2020). Analisa Spasial Clustering Zonasi Rawan Bencana Tanah Longsor Wilayah Bogor Selatan Berbasis Web-Gis. *Jurnal Inovatif: Inovasi Teknologi Informasi Dan Informatika*, 3(1), 61. <https://doi.org/10.32832/inova-tif.v3i1.4063>
- Tahun, K. K., Sadukh, J. J. P., Suluh, D. G., Rahmawaty, E., & Singga, S. (2021). Analisis Spasial Kejadian Demam Berdarah Dengue (DBD) Berdasarkan Kepadatan Penduduk dan Luas Pemukiman Di Wilker PKM Sikumana ,. 4(2), 59–63.
- Utari, R. F., Insani, F., Agustian, S., & Afriyanti, L. (2024). *Clustering Electricity Distribution Data Using the Mean Shift Algorithm Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean Shift*. 4(July), 1015–1023.
- Wihasti, I. S., Widodo, R., Pramono, D., Teknik, D., & Mada, U. G. (2025). *Tipologi dan Distribusi Spasial Bidang Tanah dalam Struktur Perkotaan : Studi Kasus Kota Denpasar*. 8(May), 268–289.